



# **The Battle of the Neighborhoods Mong Kok, Hong Kong**

**IBM Data Science Capstone Project by Chu Kai Ming**



# Mong Kok

---

- **Mong Kok** is a busy district located in Kowloon, Hong Kong. It is an area stacked with entertainment, food and beverages, local markets and big shopping malls. It is also well-known to be one of the best spots for tourists because of the neon light billboards. It was been described as the busiest district in the world by the Guinness World Records which has an extremely high density of population density of 130000/km<sup>2</sup>.

# Clustering Venues in Mong Kok

---



High rental payment for investors



Making business decisions with the help of data science



Clustering the venues according to their categories and popularity to find out the most successful business model in Mong Kok

# Data Collection

- The data source of Mong Kok is mainly from the Foursquare API.

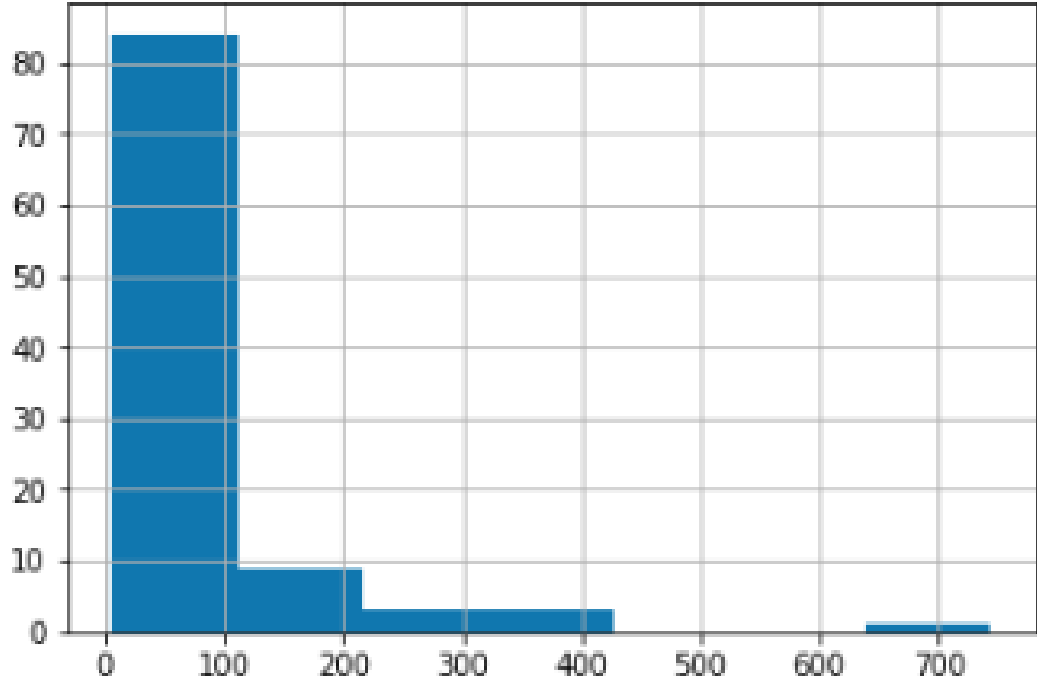
<https://developer.foursquare.com/docs/api-reference/venues/search/>



100%



## Word Cloud by the Venues' Categories



## Histogram by the Venues' Likes

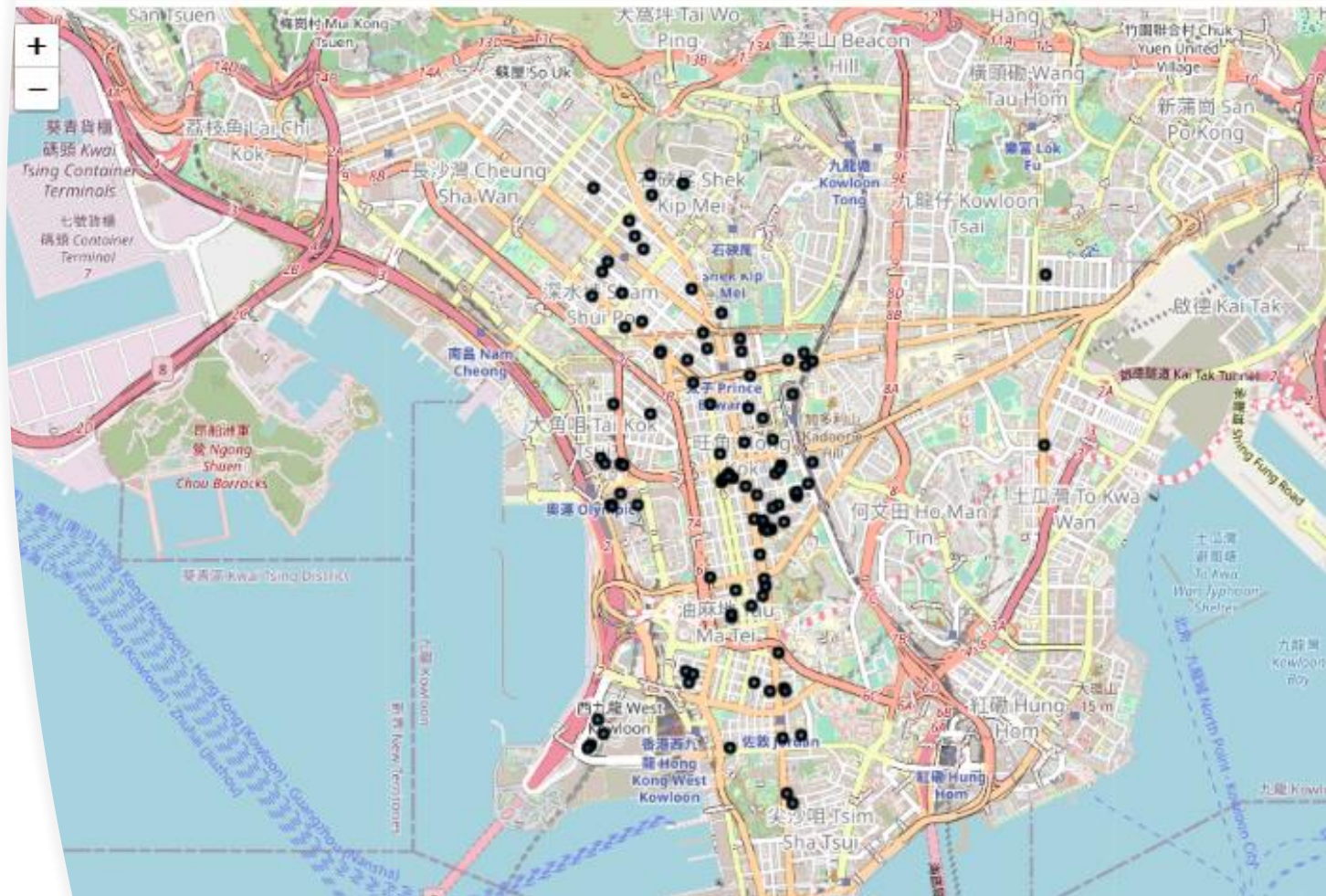


# Data Visualization (Before Analysis)

- A map generated by **Folium** visualize the location of the venues could give more information when making the data analysis decisions.

```
lium  
  
: folium.Map(location=[latitude, longitude], zoom_start=14)  
, lng in zip(mk_venues['lat'], mk_venues['lng'])):  
lium.CircleMarker(  
    [lat, lng],  
    radius=3,  
    color='black',  
    fill=True,  
    fill_color='#3186cc',  
    fill_opacity=0.7,  
    parse_html=False).add_to(map_mk)
```

mk



# Data Labeling and Re-categorization

- Labeling the venues with a 5-level tier list.
- Recategorize restaurants that selling local food into a new category called "hkfood".
- Recategorize different kind of bars as "drinks".

Out[146]:

	name	id	categories	lat	lng	likes	Tier	label	new_cat
0	Cordis, Hong Kong (香港康得思酒店)	4b0588ccf964a5207eda22e3	Hotel	22.318175	114.168112	191	5	5	Hotel
1	T. A. P. - The Ale Project	54819bb2498e42756eb3fe49	Beer Bar	22.317495	114.172610	182	5	3	drinks
2	Kam Wah Café (金華冰廳)	4bb85b883db7b7133340219a	Cha Chaan Teng	22.322275	114.169755	392	5	7	hkfood
3	Green Common The FOREST	59a28fa993bd63511b9cd8cd	Vegetarian / Vegan Restaurant	22.319138	114.171755	13	2	1	Vegetarian / Vegan Restaurant
4	Chuan Spa (「川」水療中心)	4bb5dd2aef159c74c01a75f7	Spa	22.318213	114.168099	14	2	1	Spa
5	Black Sugar Coffee	56dbd932498edb85546c912f	Coffee Shop	22.319294	114.173588	49	4	8	Coffee Shop
6	White Noise Records	4c672bd2d3899c7464a5002a	Record Shop	22.322509	114.167452	22	3	0	Record Shop
7	Ming Court (明閣)	4bbffe322a89ef3bb107f088	Cantonese Restaurant	22.318420	114.168253	74	4	2	hkfood

# One Hot Encoding

- Apply one hot encoding on the new categories and tier list representing their popularity.

```
In [131]: mk_onehot = pd.get_dummies(mk_venues[['new_cat', 'Tier']], prefix="", prefix_sep="")
mk_onehot['Name'] = mk_venues['name']
fixed_columns = [mk_onehot.columns[-1]] + list(mk_onehot.columns[:-1])
mk_onehot = mk_onehot[fixed_columns]
mk_onehot.head()
```

Out[131]:

[illegible]



```
For n_clusters=2, The Silhouette Coefficient is 0.13784926415683088
For n_clusters=3, The Silhouette Coefficient is 0.19842058771096205
For n_clusters=4, The Silhouette Coefficient is 0.2625726720065478
For n_clusters=5, The Silhouette Coefficient is 0.32280036546879187
For n_clusters=6, The Silhouette Coefficient is 0.29000516145206295
For n_clusters=7, The Silhouette Coefficient is 0.29030517914198223
For n_clusters=8, The Silhouette Coefficient is 0.27281817632401223
For n_clusters=9, The Silhouette Coefficient is 0.35359198754254445
For n_clusters=10, The Silhouette Coefficient is 0.24095927601809936
For n_clusters=11, The Silhouette Coefficient is 0.22962594268476605
```

The most effective cluster number is 9

## The Silhouette Coefficient

The k-clusters is set to 9 and the k means labels are generated.

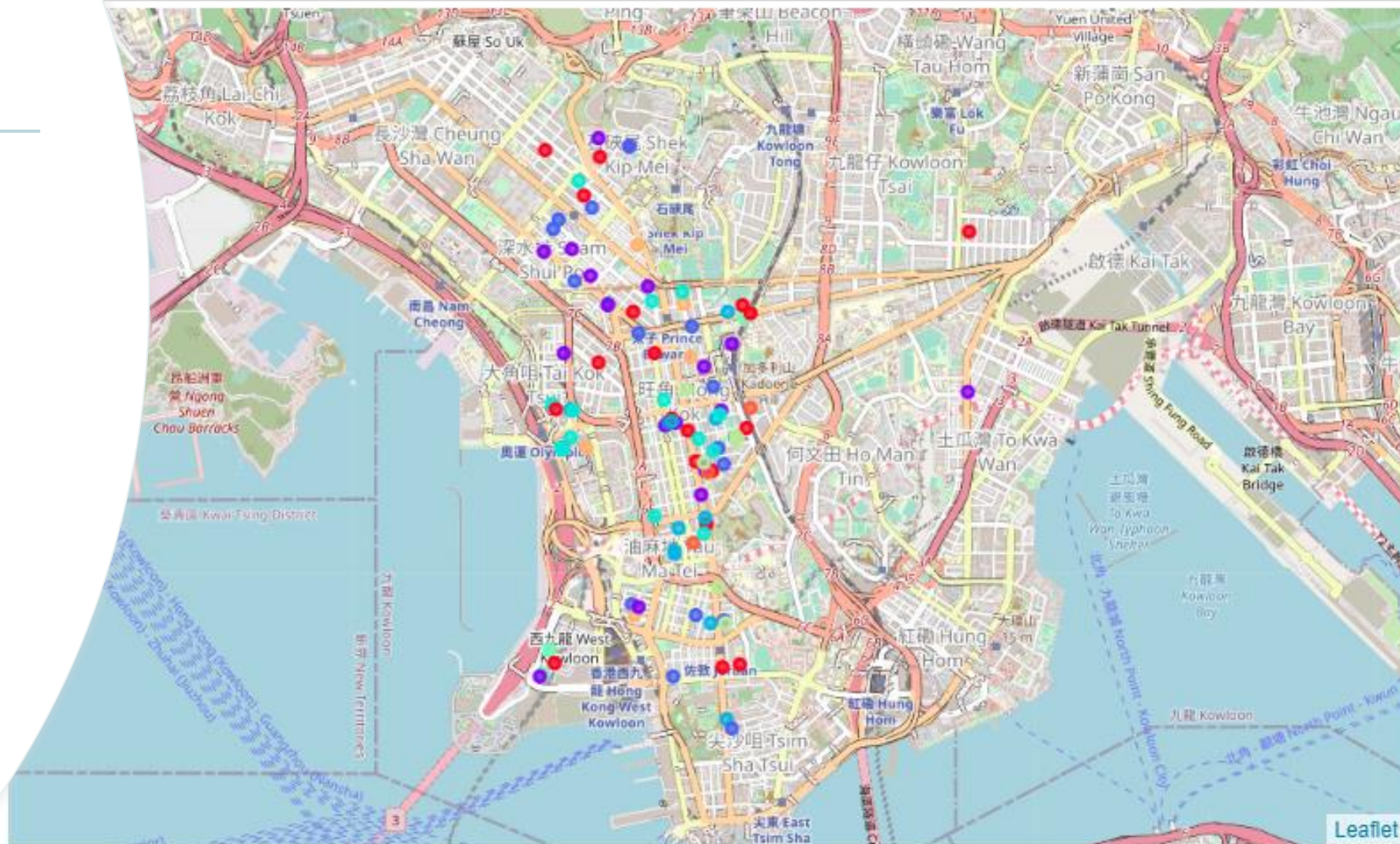
```

markers_colors = []
for lat, lon, poi, cluster in zip(mk_venues['lat'], mk_venues['lng'], mk_venues['name'], mk_venues['label']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=4,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

```

## Results

- The map is generated by **Folium** with different colors according to the labels they get from different clustering groups.





# Cluster Results

Here are 4 cluster results out of 9 and the details could be checked out in the Jupyter notebook.

<https://github.com/siuyuk/IBM-Data-Sci-Capstone-Project/blob/master/mongkokse%26cluster.ipynb>

```
In [171]: mk_venues.loc[mk_venues['label']==2]
Out[171]:
```

	name	id	categories	lat	lng	likes	tier	label	new_cat
17	Chuan Court (傳記)	4a9b5322a0f9e36a1917608	Cantonese Restaurant	22.318420	114.163253	74	4	2	Inflood
21	Urban Coffee Roaster	54ca8f952d86a2489237746c	Cafe	22.325490	114.164429	42	4	2	Cafe
22	Fai Jai (肥仔小食店)	4cb15765aaf156ab408b6954	Snack Place	22.319933	114.171859	50	4	2	Snack Place
27	Ming Yuen Restaurant (明園酒家)	4c075261281e0c6668a1e1d	Cha Chaan Teng	22.316680	114.171562	28	4	2	Inflood
28	Yuhe Tai (渝合泰)	43260308060a0e07709cf23a	Ramen Restaurant	22.320562	114.171210	40	4	2	Ramen Restaurant
39	Qi Man Seng Kitchen (義美生堂)	4b276737a0b0530a724b45	Hong Kong Restaurant	22.326787	114.162321	40	4	2	Inflood
43	Thru Restaurant (透樓)	4b164c0b64a205b4248a3	Steakhouse	22.324120	114.160842	33	4	2	Steakhouse
48	Yau Yuen Xian Jai (源園鮮)	4d32a09759174593a9324d9	Sham Lei Restaurant	22.307620	114.160510	60	4	2	Sham Lei Restaurant
60	合益華小食	4c20841697d0947a0a33ea	Snack Place	22.330523	114.161229	69	4	2	Snack Place
62	Sun Hong Yuen (源園園)	4bc47f2c4c0c94eaf9921	Cha Chaan Teng	22.320862	114.160660	74	4	2	Inflood
63	W Chan Shum Tung (源園園)	47f0b917a0b0a0a0a0a0a0	Dumpling Restaurant	22.323754	114.164416	32	4	2	Dumpling Restaurant
68	Kung Wo Dou Ban Ching (公和堂)	4c8b3a9d681017f6a4836	Snack Place	22.321176	114.163371	63	4	2	Snack Place
72	Yau Yuen Xian Jai (源園鮮)	4b5d4e4b040530a024848	Hotel	22.306650	114.171620	29	4	2	Hotel
74	Manalimania	4c29c31985d134a0b0b07	Indian Restaurant	22.305959	114.170081	32	4	2	Indian Restaurant
89	Colour Brown	9d4c47708a79330a206	Cafe	22.303279	114.160626	32	4	2	Cafe
91	Jockey Club Creative Arts Centre (香港賽馬會創意中心)	4b0588a06a45237f023a3	Art Gallery	22.314837	114.165773	35	4	2	Art Gallery
87	Max Place 2	501eac022a0b1eaf0e10204	Shopping Mall	22.300233	114.172388	53	4	2	Shopping Mall

```
In [170]: mk_venues.loc[mk_venues['label']==1]
Out[170]:
```

	name	id	categories	lat	lng	likes	tier	label	new_cat
3	Shan Chuan Tea House (山園茶樓)	4b2d0a0f03a0131b0a0b0a	Vegetarian / Vegan Restaurant	22.319138	114.171765	13	2	1	Vegetarian / Vegan Restaurant
4	Chuan Court (傳記)	4a9b5322a0f9e36a1917608	Snack	22.318420	114.163253	74	2	1	Snack
9	Superman Toys	4b7a0a0a0a0a0a0a0a0a0a	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
10	Superman Toys	4b7a0a0a0a0a0a0a0a0a0a	Food & Drink Shop	22.318420	114.170679	17	2	1	Food & Drink Shop
11	Woo & Spencer Food	4b10a20f0a0a0a0a0a0a0a	Cooking Store	22.318420	114.161517	13	2	1	Cooking Store
12	Woo & Spencer Food	4b10a20f0a0a0a0a0a0a0a	Cooking Store	22.318420	114.161517	13	2	1	Cooking Store
24	The Horizon Restaurant (海平線)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
25	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
26	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
27	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
28	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
29	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
30	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
31	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
32	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
33	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
34	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
35	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
36	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
37	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
38	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
39	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
40	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
41	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
42	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
43	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
44	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
45	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
46	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
47	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
48	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
49	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
50	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
51	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
52	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
53	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
54	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
55	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
56	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
57	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
58	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
59	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
60	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
61	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
62	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
63	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
64	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
65	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
66	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
67	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
68	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
69	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
70	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
71	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
72	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
73	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
74	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
75	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
76	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
77	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
78	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
79	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
80	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
81	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
82	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
83	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
84	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
85	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
86	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
87	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
88	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
89	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
90	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
91	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
92	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
93	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
94	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
95	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
96	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
97	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
98	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
99	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store
100	Bound by History (歷史的束縛)	4b5d4e4b040530a024848	Toy / Game Store	22.318420	114.170679	17	2	1	Toy / Game Store

#	name	id	categories	lat	lng	likes	tier	label	new_cat
6	White Noise Records	4a072b0d0309b1b1a0a0b0a	Record Shop	22.323590	114.161462	10	3	0	Record Shop
9	Parade Records (樂隊)	57565a0a0a0a0a0a0a0a0a	Dumpling Restaurant	22.317951	114.169590	21	3	0	Dumpling Restaurant
12	Wolf Club Bar	56326b075406b0a0b00000	Beer Bar	22.319030	114.173396	22	3	0	Beer Bar
14	Hot Day, Sweet Day	42635b91605406b05675303	Taiwanese Restaurant	22.316058	114.170127	22	3	0	Taiwanese Restaurant
20	King of Coconut (椰子王)	4440199e1a672ba07830a07	Juice Bar	22.315529	114.171991	22	3	0	Juice Bar
21	Pepper Lunch	404a13209a46a02952a02a2	BBQ Joint	22.316000	114.168352	26	3	0	BBQ Joint
30	Chi Square (奇平方)	48a427a17a060a0a0a0a0a	Bar	22.312462	114.170772	20	3	0	Bar
31	The Brew Lab	405a0a0a0a0a0a0a0a0a0a	Coffee Shop	22.319328	114.171410	24	3	0	Coffee Shop
44	Hay Fever Club & Gigs	4a9f7013018727a0a0a0a0	Bar	22.325174	114.170358	25	3	0	Bar
49	Lon Rock (浪客)	00940a0a0a0a0a0a0a0a0a	Bar	22.324912	114.170040	24	3	0	Bar
54	純純咖啡屋 (Chun Chun Coffee House)	40a4290a0321a0a0a0a0a0	Noodle House	22.319187	114.163565	22	3	0	Noodle House
64	Man Ka Teat Noodle (文記茶樓)	40a473101010a0a0a0a0a0	Noodle House	22.323989	114.162849	24	3	0	Noodle House
70	神神堂食店	42a9e33010023210a0a0a0	Dim Sum Restaurant	22.324914	114.165054	25	3	0	Dim Sum Restaurant
84	Calli REVEL	58b31101a0a0a0a0a0a0a0	Cafe	22.304831	114.172590	24	3	0	Cafe
90	Yee Yee Yee (Yee Yee Yee 永和) 永和豆漿大王	51ab339a0a0a0a0a0a0a0a	Breakfast Restaurant	22.334195	114.163885	23	3	0	Breakfast Restaurant
94	China Tuna Restaurant	54c47a0a0a0a0a0a0a0a0a	Japanese Restaurant	22.316357	114.163565	24	3	0	Japanese Restaurant
95	Yee Yee Yee (Yee Yee Yee 永和) 永和豆漿大王	51ab339a0a0a0a0a0a0a0a	Breakfast Restaurant	22.334214	114.163885	24	3	0	Breakfast Restaurant
98	Kung Yui Don Chuan (龍鳳燒臘)	4a9b2810252a0a0a0a0a0a	Dessert Shop	22.32078	114.165843	23	3	0	Dessert Shop
99	Genki Soba (元氣蕎麥粉)	4a3980310a0a0a0a0a0a0a	Sushi Restaurant	22.303047	114.171831	21	3	0	Sushi Restaurant

# Conclusions and Improvements

---



The data source of this project is not big and complete enough to make it to be great.



The target audience of this report should get more information about the venues that should be recommended in different clusters.



If I could have some improvements for the analysis, I would try to find more data or change the attributes for one hot encoding and clustering number to finalize the results better.